



Junta de Andalucía

Consejería de Salud y Consumo  
Servicio de Estadísticas sanitarias  
Unidad Estadística y Cartográfica

# MEMORIA TÉCNICA DE LA ACTIVIDAD

## “BASE POBLACIONAL DE SALUD”

0. IDENTIFICACIÓN.....	3
1. INTRODUCCIÓN.....	4
2. ÁMBITO DE ESTUDIO.....	8
3. RECOGIDA O CAPTURA DE DATOS.....	9
4. FLUJO O PROCESO DE TRABAJO.....	10
5. PLAN DE DIFUSIÓN.....	13
6. CALIDAD.....	14





## 0. IDENTIFICACIÓN

- **Código y denominación de la actividad:** 04.02.15 Base Poblacional de Salud (BPS)
- **Organismo responsable:** Consejería de Salud y Consumo
- **Unidad ejecutora:** Servicio de Producto Sanitario. Subdirección Técnica Asesora de Gestión de la Información. Servicio Andaluz de Salud
- **Organismos colaboradores y convenio:**

**Convenio de colaboración** entre la Consejería de Salud de la Junta de Andalucía y el Instituto de Estadística y Cartografía de Andalucía para el desarrollo de actividades estadísticas y cartográficas contenidas en el Plan Estadístico y Cartográfico de Andalucía.



## 1. INTRODUCCIÓN

- **Objetivos:**

- Identificar las necesidades de atención sanitaria de la población en su conjunto y de subgrupos poblacionales.
- Apoyar el desarrollo de las actividades de planificación sanitaria.
- Apoyar el desarrollo de las actividades en el ámbito de la salud pública.
- Apoyar la gestión sanitaria así como el uso y mejora de la distribución de recursos sanitarios.
- Anticiparse a las necesidades sanitarias futuras.
- Configurar biografías sanitarias que permitan realizar estudios longitudinales sobre la salud y la utilización de servicios sanitarios.
- Desarrollar un sistema de información que conecte datos procedentes de diferentes sistemas de información del SSPA y de otras fuentes externas, para su uso con fines estadísticos.
- Disponer de una infraestructura de datos para la investigación.

- **Marco conceptual:**

La digitalización de los servicios sanitarios está permitiendo el almacenamiento de una gran cantidad de datos referentes a la salud de la población y a su interacción con dichos servicios. Sin embargo, la complejidad de estos sistemas de información (historia clínica digital y otros) hace que la explotación de datos se realice de forma aislada por cada módulo, lo que conlleva la creación de silos independientes con la consecuente aparición de duplicidades e infrautilización de la información, afectando a su calidad y a los costes. Se precisa, por tanto, de la conexión entre diferentes bases de datos para que estos análisis permitan informar de forma adecuada a los diversos responsables de la toma de decisiones.

Dentro del concepto de Sistema Sanitario Público de Andalucía (SSPA) se incluyen todos los centros que prestan asistencia sanitaria, de cualquier tipo, a la población, tanto los gestionados directamente por el Servicio Andaluz de Salud (SAS) como aquellos gestionados por Agencias Públicas. Para la prestación de esta asistencia se utilizan diferentes sistemas de información que interrelacionan entre ellos.

La Base Poblacional de Salud es un Sistema de Información Sanitaria (SIS) de base poblacional que recoge datos de toda la población incluida en la Base de Datos de Usuarios del SSPA (BDU).

El núcleo de la información lo proporciona BDU, aportando datos demográficos que se complementa con la información diagnóstica y de utilización de los recursos sanitarios proveniente de los distintos ámbitos asistenciales de atención primaria y especializada.



Entre las fuentes administrativo-clínicas que la alimentan destacan la historia clínica digital y los diferentes Conjuntos Mínimos Básicos de Datos (CMBD) hospitalarios (hospitalización, hospitales de día y urgencias). En cuanto al uso de recursos, incluye las consultas de Atención Primaria (AP) y Atención Hospitalaria (AH), las urgencias, los distintos procesos de atención hospitalaria y el consumo de farmacia, de sesiones de diálisis y de transporte sanitario.

Estos datos, que constituyen una auténtica biografía sanitaria de los individuos, permitirán realizar estudios longitudinales, estudiar la historia natural de las enfermedades en el conjunto de la población o en determinados subgrupos, analizar la incidencia de patologías o realizar proyecciones sobre el estado de la salud y las necesidades de recursos, así como análisis transversales y estratificación de la población para orientar la prestación de los servicios o la aproximación a la prevalencia de distintas patologías.

En estos momentos dispone de datos de 13 millones de usuarios (2001 – 2019), con un total de 515 millones de diagnósticos (88% de ellos obtenidos mediante un proceso de codificación automática). Estos diagnósticos se distribuyen en 408 millones procedentes de AP (desde 2001), 51 millones de altas de hospitalización y 4,5 millones de altas de cirugía mayor ambulatoria (desde 2001), 48,5 millones de urgencias AH (desde 2009) y 3 millones de consultas externas de salud mental (desde 2014). También contiene 2,7 millones de valoraciones funcionales y cognitivas (desde 2014) y 12 millones de diagnósticos de enfermería (desde 2004). Además cuenta con 699 millones de resultados de pruebas analíticas (desde 2017).

Respecto a los datos administrativos y clínicos, las bases que se integran en la BPS son:

- La Base de Datos de Usuarios (BDU), donde se asigna un número de identificación único para el sistema sanitario (NUHSA) y se relaciona a cada usuario con una referencia asistencial asignándolo a un cupo médico, localizado en un centro de salud. Estos centros de salud se agrupan en Distritos Sanitarios, que tienen a su vez un área hospitalaria de referencia.
- Historia de salud electrónica (DIRAYA)
- Conjunto Mínimo Básico de Datos (CMBD): Hospitalización, Cirugía mayor ambulatoria (CMA), Urgencias hospitalarias y Hospital de día médico (HDM)
- La Base de Citas (Citación), donde se asignan las citas a las consultas externas de los dos niveles asistenciales, Atención Primaria (AP) y Atención Hospitalaria (AH)
- Gestión de prestaciones farmacéuticas
- Registro de vacunaciones de Andalucía
- Sistema de información de pacientes renales

Respecto a las acciones que se realizan para integrar las diferentes bases de datos, se pueden definir:

- Fusión de registros: proceso por el cual se establecen relaciones de correspondencia entre informaciones de un mismo individuo que proceden de fuentes de información diferentes.



- Enlace de registros: proceso por el que se determina que dos registros pertenecen al mismo individuo.

- **Marco jurídico:**

- La normativa que regula la historia clínica está contenida en la ley 14/1986, de 25 de abril, General de Sanidad (BOE núm. 102 de 29/04/1986) y en la ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica (BOE núm. 274 de 15/11/2002).
- La creación de la Base Poblacional de Salud se realiza mediante [resolución 0068/18 de la Dirección Gerencia del Servicio Andaluz de Salud, con fecha 18/05/2018](#).
- Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) y Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
- Directiva Europea Inspire (Directiva 2007/2/CE, Infrastructure for Spatial Information in Europe) que establece las reglas generales obligatorias para el establecimiento de una Infraestructura de Información Espacial en la Comunidad Europea basada en las Infraestructuras de los Estados miembros.
- Ley 12/1989, de 9 de mayo, de la Función Estadística Pública.
- Ley 4/1989, de 12 de diciembre, de Estadística de la Comunidad Autónoma de Andalucía.
- Ley 9/2023 de 25 de septiembre, por la que se aprueba el Plan Estadístico y Cartográfico de Andalucía 2023-2029 y sus programas estadísticos y cartográficos de desarrollo.

- **Antecedentes:** Hay antecedentes del uso secundario de los datos clínicos y de interacciones de los pacientes con los servicios de salud. Estos usos permiten reconstruir la biografía sanitaria de las personas, abriendo posibilidades de análisis longitudinal para la planificación, gestión e investigación sanitarias. Además, puede ayudar a la toma de decisiones en los ámbitos estratégico (planes y estrategias), mesogestión (contratos programa) y microgestión (gestión clínica) con la facilidad de uso de los informes con indicadores predefinidos y la autonomía del análisis dinámico de datos. Dado que el SSPA presta atención a la mayoría de la población andaluza, podemos considerar que la cobertura del sistema es poblacional.

En cuanto a actividad estadística oficial se incorporó por primera vez en el Programa Estadístico y Cartográfico 2017.

- **Justificación y utilidad:** Con este proyecto se está desarrollando una nueva fuente de información estadística a partir de la integración y reutilización de información ya recopilada en registros estadísticos y administrativos sanitarios preexistentes. La integración de fuentes es un medio



eficiente para mejorar la calidad y precisión de las fuentes preexistentes y un medio de generar nuevas series y productos estadísticos a un coste mucho menor que el que supondría desarrollar operaciones de campo o crear nuevos circuitos para la recogida directa de la información.

Es una manera de gestionar la información que ya existe conectando distintas fuentes de información que recogen datos de las personas, permitiendo reconstruir el recorrido vital de las personas en cuanto a su salud y la atención a la misma.

- **Restricciones y alternativas:**

Las mayores restricciones vienen dadas por la calidad de la cumplimentación de la historia clínica e informe de alta, relacionada con la indexación y codificación de los códigos clínicos y algunas de las variables estudiadas (código postal, ocupación, hospitales de procedencia y traslado, y unidad de gestión clínica al alta).

Para subsanar estas restricciones, el Servicio de Producto Sanitario realiza:

- Asesoramiento continuo a personal documentalista de las instituciones del SSPA.
- Cursos de formación en codificación mediante la CIE-10-ES dirigidos a los profesionales encargados de la codificación en los hospitales del SSPA.
- Auditorías sistemáticas con el objetivo de estudiar la fiabilidad y calidad de este sistema de información y de conocer las principales características de los servicios de documentación. Los aspectos básicos sobre los que versan estas auditorías hacen referencia a la disponibilidad y calidad de la historia clínica e informe de alta, los errores de las variables administrativas (sexo, fecha de nacimiento o edad, NUHSA, fecha de ingreso, fecha de alta, tipo de alta, financiación y residencia), los errores de indexación y la exhaustividad y exactitud en la codificación de diagnósticos y procedimientos.

Además, para conseguir una mayor calidad de los datos se ha implantado un programa de validación de datos que comprueba la unicidad, la coherencia y la exhaustividad de los mismos permitiendo a los hospitales su corrección.

- **Comparabilidad territorial:** No se conocen antecedentes en otros ámbitos ni a nivel nacional.



## 2. ÁMBITO DE ESTUDIO

- **Objeto de estudio:** conjunto de apuntes registrales personales que el sistema sanitario recopila de manera sistemática en diversas actividades y que informan sobre la ocurrencia de eventos o estados de índole sanitario: acceso a consultas, vacunaciones, pruebas diagnósticas, hospitalizaciones, etc.
- **Resolución, escala o desagregación del objeto de estudio:**

Puntual. A partir de ahí, para su difusión, agregación de datos por los distintos ámbitos sanitarios.

- **Fenómenos o variables:**
  - Características demográficas y sociales:
    - Sexo
    - Fecha de nacimiento
    - Fecha de alta en el SSPA
    - Fecha de baja en el SSPA (Fecha de baja por defunción)
    - Vive en Zona de Transformación Social
    - Tramo de copago farmacéutico.
    - Lugar de residencia (Urbano, intermedio, rural)
  - Características clínicas:
    - Valoración funcional
    - Valoración cognitiva
    - Resultados pruebas diagnósticas y fecha.
    - Diagnóstico y fecha (A. primaria, CMBD hospitalización, c. externas y hospital de día).
  - Uso de recursos, fecha y duración en su caso:
    - Prescripción farmacéutica
    - Consultas (atención primaria, hospitales, urgencias), ingresos hospitalarios
    - Vacunas recibidas (sólo desde 2018)
    - Ciertas constantes (peso, talla, Tensión Arterial)



### 3. RECOGIDA O CAPTURA DE DATOS

- **Sujeto informante:** Servicio Andaluz de Salud
- **Tipología de datos a suministrar:** Los datos primarios que se van a obtener son datos de personas físicas.
- **Periodicidad:** Se realiza una carga de datos anual.
- **Método de obtención:**

Esta actividad se basa en la explotación de fuentes administrativas y otra información sanitaria.

#### Denominación oficial de las fuentes administrativas:

- Base de datos de Usuarios del SSPA (BDU)
- Conjunto Mínimo Básico de Datos de Andalucía (CMBD)
- Historia de Salud electrónica (Diraya)
- Gestión de prestaciones farmacéuticas
- Registro de vacunaciones de Andalucía
- Registro de Población de Andalucía

Al tratarse de un sistema basado en la historia clínica, la recogida directa de la información la realizan profesionales asistenciales de manera continua en los diferentes encuentros con cada paciente, como consecuencia del desarrollo de las funciones asistenciales atribuidas al Servicio Andaluz de Salud. Se realiza una recopilación de información de los diferentes sistemas anteriormente descritos, incorporándose en la BPS de manera mensual.



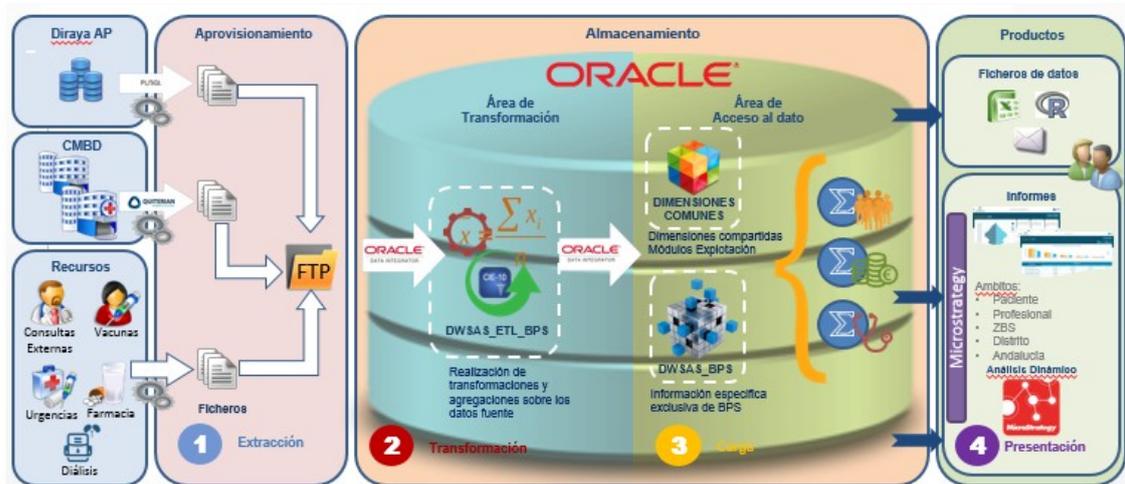
#### 4. FLUJO O PROCESO DE TRABAJO

- **Preparación y tratamiento base de la información:**

La idea marco es la reconstrucción de las secuencias sociodemográficas y sanitarias de los individuos con objeto de posibilitar su explotación estadística. Para ello se necesita disponer de un sistema que permita llegar a reconstruir las líneas de vida individuales. Disponer de este sistema permite la investigación estadística de los distintos tipos de episodios que componen la historia de salud de los individuos, como las interrelaciones que se producen entre eventos. Este esquema conceptual permite la generación de nuevas variables no directamente recogidas en un formulario estadístico o administrativo, sino producidas por el proceso de integración de la información, o las interrelaciones entre distintos tipos de eventos.

En los procesos ETL (Extract, Transform and Load) los datos se extraen de sistemas operacionales o de ficheros externos, se depositan en un servidor FTP y se cargan en la Base de Datos Oracle, tras controles de recepción y detección de inconsistencias. En el proceso de transformación, los datos se enriquecen para facilitar su explotación, realizándose operaciones de codificación de diagnósticos, aplicación de reglas de negocio, agrupación de casos (Adjusted Clinical Groups ACG y Grupos de Morbilidad Ajustados GMA) y cálculo de indicadores.

La arquitectura del sistema de información se muestra en la siguiente figura:



#### Mecanismos desarrollados para comprobar su fiabilidad

La integración de múltiples fuentes con abundancia de información redundante permite comprobar la calidad de la información recopilada por cada una de las fuentes primarias. La resolución de las discrepancias e incoherencias es un requisito imprescindible para poder realizar adecuadas explotaciones estadísticas. Con cada nueva carga de información se realizan sistemáticas comprobaciones de la coherencia de la información almacenada con la nueva información incorporada. Este sistema supone de



hecho una permanente revisión de la calidad y fiabilidad de la información almacenada conforme se van obteniendo sucesivas observaciones de un individuo concreto.

El criterio más restrictivo pasa por la coincidencia en todas las fuentes de todos los datos en un conjunto de identificadores estrictos. El enlace conseguido de esta forma establece la correspondencia entre fuentes por encima del 80% de los individuos registrados. Para el resto se busca establecer un grado de correspondencia que si bien no sea perfecta, exista un alto grado de seguridad de que ambos registros son los mismos. Este grupo de enlace incluye necesariamente un componente probabilístico.

Para asegurar la fiabilidad del sistema se toma una pequeña muestra de estos enlaces y se revisan manualmente para estimar si hay enlaces asimilados como verdaderos sin evidencia clara de que lo sean. Estos posibles enlaces erróneos son denominados falsos positivos. El análisis de estos pares revela que la tasa de falsos positivos es reducida, lo que garantiza que la información emparejada es bastante fiable para su tratamiento estadístico.

Una vez obtenido los enlaces, hay que establecer un orden de prioridad para resolver los posibles conflictos entre la información reportada por las distintas fuentes.

Un paso fundamental para asegurar la coherencia de la información almacenada en la Base Poblacional de Salud es la construcción de los episodios a partir de la información cruda de eventos y observaciones almacenada en las fuentes. Los episodios muestran la permanencia de un individuo en un estado determinado, como por ejemplo residiendo en una localización geográfica determinada, desde una determinada fecha que inicia el episodio hasta otra fecha en que concluye, por ejemplo si hablamos de residencia bien por defunción, migración a otro ámbito geográfico.

Los episodios se obtienen tras una depuración de los flujos de alta y baja, de manera que no se produzcan incongruencias y se cumpla una serie de restricciones, como la que un individuo sólo puede estar asociado a una única residencia en el mismo instante temporal. Para ello se realizan las oportunas depuraciones e imputaciones, como suprimir los eventos repetidos y modificar, en su caso, las fechas de inicio o fin de episodio para garantizar la consistencia espacio-temporal en la secuencia de eventos.

Esta actividad estadística sirve de base para proporcionar información a la unidad ejecutora y otras unidades y actividades estadísticas para que éstas procesen y analicen estos datos posteriormente. En particular, existen dos actividades estadísticas directamente relacionadas con la Base Poblacional de Salud: La Base de Datos de Usuarios y El Conjunto Mínimo Básico de Datos al alta. Ambas también se encargan de realizar los análisis y las actividades de difusión directamente asociadas con esta fuente.

- **Garantía del secreto estadístico y protección de datos personales.**

La Base poblacional de salud se encuentra almacenada en el Sistema Informático del SAS y es gestionado por personal propio de este organismo, y por lo tanto sometido a los estándares de seguridad y control de acceso claramente estipulados por el protocolo de seguridad de este organismo. En todo caso su gestión respeta escrupulosamente el marco normativo del secreto



estadístico que impide la utilización y transferencia de la información almacenada en el sistema para fines no estadísticos. Igualmente la Base poblacional de Salud se rigen por la normas y medidas de seguridad recogidas en Ley Orgánica de Protección de Datos de Carácter Personal y garantía de derechos digitales.

- **Codificación, estándares, nomenclaturas y clasificaciones utilizadas:**

Los diagnósticos clínicos son codificados por documentalistas clínicos para el CMBD de hospitalización, de CMA y de HDM y mediante un codificador automático de desarrollo propio, para urgencias hospitalarias e historia clínica electrónica. Para la codificación se utilizó la modificación clínica de la Clasificación Internacional de Enfermedades (CIE), tanto la novena versión española (CIE-9-MC) de 2014 como primeras versiones CIE-10-ES. También se utiliza la clasificación NANDA para los diagnósticos de enfermería y los códigos del catálogo de pruebas diagnósticas.

- **Mantenimiento, conservación y actualización:** La información descargada anualmente se guarda en carpetas bien identificadas y se almacena en directorios perfectamente organizados y gestionados con formatos y nomenclatura establecida.



## 5. PLAN DE DIFUSIÓN

- **Producto 1:** Base Poblacional de Salud
- **Tipo de resultados y formatos:** Tablas y gráficos, descargables en formato PDF, Excel, PC-Axis, CSV, Ods, JSON.
- **Periodicidad:** Anual
- **Usuarios:** actualmente no se realizan encuestas de satisfacción y ni de la calidad percibida por los usuarios. Por lo que no se aplica al proceso de generación de la información el feedback del usuario.
- **Producto 2:** Memoria estadística de la Consejería.
- **Tipo de resultados y formatos:** Tablas y gráficos, descargables en formato PDF, Excel, PC-Axis, CSV, Ods, JSON.
- **Periodicidad:** Anual
- **Usuarios:** actualmente no se realizan encuestas de satisfacción y ni de la calidad percibida por los usuarios. Por lo que no se aplica al proceso de generación de la información el feedback del usuario.



## 6. CALIDAD

### **Reproducibilidad del proceso:**

Tanto la subdirección como la empresa encargada de la extracción de los datos de los sistemas operacionales correspondientes mantienen la documentación de los diferentes procesos y esquemas de la base de datos, de manera que estos procesos de ETL son reproducibles en cualquier momento y por cualquier operador con el conocimiento técnico necesario. También se documentan los cambios que se puedan producir en los diferentes sistemas de codificación, de manera que pueden recomponerse los procesos. Esta documentación se encuentra en carpetas organizadas en el servidor corporativo, del que se hace copia de seguridad, y a las que tienen acceso los integrantes de la subdirección.

### **Oportunidad y puntualidad:**

**Oportunidad:** Los datos se publican en el segundo trimestre del año siguiente al que se refieren los datos.

**Puntualidad:** los datos se divulgan en las fechas anunciadas en el calendario de publicaciones.

• **Respecto a los procesos:** se sigue un flujo de trabajo establecido y estandarizado descrito en los apartados anteriores lo cual permite que la calidad de la información generada se robusta y fiable.

### • **Respecto a los resultados:**

◦ *Relevancia y utilidad:* la información generada es relevante y de utilidad debido a que abarca la totalidad de la población objeto de estudio.

◦ *Precisión y confiabilidad:* la información es precisa en cuanto que las variables diseñadas ofrecen la información real y total del fenómeno medido, y confiable en cuanto a que la información se genera según lo establecido en los manuales actualizados por el Ministerio a través de un proceso planificado y estandarizado.

◦ *Nivel de estandarización o conformidad:* los datos y metadatos han sido obtenidos aplicando los principios del “Código de Buenas Prácticas de las Estadísticas Europeas”. Para el proceso de conformación de la información se utiliza el programa estadístico R.

◦ *Esquema de calidad:* no se sigue ningún estándar.